

Thresholds for epidemic spreading in networks

Claudio Castellano¹ and Romualdo Pastor-Satorras²

¹*Istituto dei Sistemi Complessi (CNR-ISC), UOS Sapienza and Dip. di Fisica,
"Sapienza" Università di Roma, P.le A. Moro 2, I-00185 Roma, Italy*

²*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord B4, 08034 Barcelona, Spain*

We study the threshold of epidemic models in quenched networks with degree distribution given by a power-law. For the susceptible-infected-susceptible model the activity threshold λ_c vanishes in the large size limit on any network whose maximum degree k_{max} diverges with the system size, at odds with heterogeneous mean-field (HMF) theory. The vanishing of the threshold has nothing to do with the scale-free nature of the network but stems instead from the largest hub in the system being active for any spreading rate $\lambda > 1/\sqrt{k_{max}}$ and playing the role of a self-sustained source that spreads the infection to the rest of the system. The susceptible-infected-removed model displays instead agreement with HMF theory and a finite threshold for scale-rich networks. We conjecture that on quenched scale-rich networks the threshold of generic epidemic models is vanishing or finite depending on the presence or absence of a steady state.

PACS numbers: 89.75.Hc, 05.70.Ln, 87.23.Ge, 89.75.Da

The heterogeneous pattern of a network can have dramatic effects on the behavior of dynamical processes running on top of it [1], in particular when the distribution of the number k of contacts (the degree of an element or vertex) exhibits long tails, as expressed by a power-law degree probability with the asymptotic form $P(k) \sim k^{-\gamma}$ [2]. An example that has attracted a great interest due to its practical real-world implications is the modeling of epidemic spreading on contact networks [3]. The simplest of these models is the SIS model [4], in which each vertex (individual) can be in one of two states, either susceptible, or infected. Susceptibles become infected by contact with infected individuals, with a rate proportional to the number of infected contacts times a given spreading rate λ . Infected individuals on the other hand become healthy again with a rate that can be set arbitrarily equal to unity. The model allows thus individuals to contract the infection time and again, leading, in the infinite network size limit, to a sustained infected steady state for values of λ larger than an epidemic threshold λ_c . In the SIR model [4], on the other hand, infected individuals recover (or die) and cannot change further their state. No steady state is now allowed, but a threshold still exists above which the total number of infected individuals, starting from a very small infected seed, reaches a finite fraction of the network. The analysis of these and other models [1], performed via a mean-field theory modified to take into account the heterogeneity of the network substrate [5, 6], led to the far-reaching conclusion that topological fluctuations, as measured by the second moment of the degree distribution $\langle k^2 \rangle$, can have profound effects in many types of dynamics [1, 6]. Thus, for example, in the SIS model, the threshold takes the values, at the mean-field level $\lambda_c = \langle k \rangle / \langle k^2 \rangle$. For a long-tailed degree distribution with power law form, the second moment diverges for $\gamma \leq 3$, and one obtains the remarkable result of a vanishing epidemic threshold in the thermo-

dynamic limit. These results have led to the widespread belief in the distinction between *scale-free* networks with $\gamma \leq 3$, where topology is highly relevant, and *scale-rich* networks with $\gamma > 3$, where dynamical processes exhibit an essentially homogeneous mean-field behavior.

In this Letter, building on some results previously reported, we present evidence that this belief is not correct for the SIS model on quenched networks (i.e. networks whose adjacency matrix is fixed in time) and that the scale-free nature of the contact pattern has no crucial effect on the value of the epidemic threshold. We investigate the physical origin of this result, its validity for generic network structures and its consequences. On the other hand we show that for the SIR model the picture is different, a zero threshold occurring only in scale-free quenched networks.

While heterogeneous mean-field (HMF) theory is exact on annealed networks (i.e. networks whose adjacency matrix is fixed only in average [6]), results beyond HMF theory for the SIS process on quenched networks (QN) have appeared in different contexts and with various levels of rigor. Already in 2003, Wang et al. [7] argued that the epidemic threshold on an *arbitrary* undirected graph is set by the largest eigenvalue Λ_N of the adjacency matrix,

$$\lambda_c = \Lambda_N^{-1}, \quad (1)$$

see also [8, 9]. The relevance of Eq. (1) becomes evident when it is complemented with the results of Chung et al. [10], who calculated the largest eigenvalue of the adjacency matrix for a class of finite graphs with degrees distributed according to a power-law, obtaining

$$\Lambda_N = \begin{cases} c_1 \sqrt{k_c} & \sqrt{k_c} > \frac{\langle k^2 \rangle}{\langle k \rangle} \ln^2(N) \\ c_2 \frac{\langle k^2 \rangle}{\langle k \rangle} & \frac{\langle k^2 \rangle}{\langle k \rangle} > \sqrt{k_c} \ln(N) \end{cases}, \quad (2)$$

where N is the network size, k_c is network cut-off or degree of the most connected node (averaged over many

network realizations [11]), and c_i are constants of order 1. The cut-off k_c is a growing function of the network size for uncorrelated scale-free networks, taking the value $k_c \sim N^{1/2}$ for $\gamma \leq 3$ and $k_c \sim N^{1/(\gamma-1)}$ for $\gamma > 3$ [12]. For $\gamma > 3$ the ratio of the moments is finite and it is clear that the largest eigenvalue is governed by k_c . Noticeably this remains true also for $5/2 < \gamma < 3$, since in that range $\langle k^2 \rangle / \langle k \rangle \sim k_c^{3-\gamma} \ll \sqrt{k_c}$. Only for $2 < \gamma < 5/2$ the largest eigenvalue is set by the moments of the degree distribution. Combining Eqs. (1) and (2), the behavior of the threshold for the SIS model in a power-law distributed network is, for sufficiently large size,

$$\lambda_c \simeq \begin{cases} 1/\sqrt{k_c} & \gamma > 5/2 \\ \frac{\langle k \rangle}{\langle k^2 \rangle} & 2 < \gamma < 5/2 \end{cases}, \quad (3)$$

see also [13]. Since k_c grows as a function of N for any γ , the consequence of Eq. (3) is remarkable: *In any uncorrelated quenched random network with power-law distributed connectivities, the epidemic threshold for SIS goes to zero as the system size goes to infinity.* This has nothing to do with the scale-free nature of the degree distribution: It is always true as long as the cut-off k_c diverges. Remarkably the threshold goes to zero also for Erdős-Rényi graphs (although logarithmically slow), for which a formula similar to Eq. (2) exists [14]. Different approaches [13, 15, 16], have also pointed out that in the thermodynamic limit, the system is active for any $\lambda > 0$. These results, however, have gone largely unnoticed within the statistical physics community.

A first issue raised by Eq. (3) concerns the fact that, as any critical point, the epidemic threshold is well defined only in the thermodynamic limit. In a finite system, the dynamics is always doomed to fall into the healthy, absorbing state, even far above the threshold, due to stochastic fluctuations. The threshold for a finite network of size N must therefore be intended as the value separating the regime $\lambda < \lambda_c$ for which the epidemics decays exponentially fast (so that the expected survival time is of the order $\tau \sim \ln(N)$) from the regime $\lambda > \lambda_c$ where the survival time grows exponentially with N to some power, $\tau \sim e^{N^\alpha}$, with $\alpha > 0$.

To investigate the validity of these results we have performed numerical simulations of the SIS model on quenched scale-rich networks with $\gamma = 4.5$ and minimum degree $k_{min} = 3$, built using the uncorrelated configuration model [17]. In order to compare results with the predictions in Eq. (3) one must take into account that the actual maximum degree k_{max} in each network realization is a random variable, with average value $\langle k_{max} \rangle = k_c$. In particular, in the case $\gamma > 3$, one can see [18] that both the mean and the standard deviation of k_{max} scale as $k_c \sim N^{1/(\gamma-1)}$, implying that k_{max} always shows large fluctuations for different realizations of the degree sequence. Therefore, we first consider networks in which k_{max} has a fixed value, equal to the mean k_c numerically

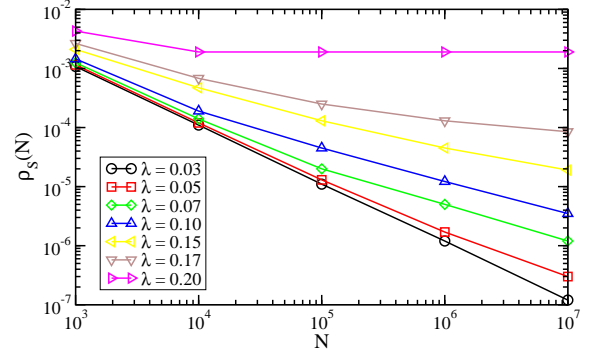


FIG. 1: Density of active sites for long times (restricted to surviving runs) in the SIS model on QN as a function of system size N , for $\gamma = 4.5$ and different values of the parameter λ . Notice that the straight line for $\lambda = 0.03$ is due to the fact that no density smaller than $1/N$ can occur.

estimated for the chosen system size N . In Fig. 1 we plot the density ρ_s , calculated only for surviving runs, as a function of N for different values of λ [19]. Should the transition occur at a fixed value of λ , ρ_s would go to a constant for $\lambda > \lambda_c$, decay exponentially for $\lambda < \lambda_c$ and as a power-law exactly at the transition. A completely different behavior is observed: for all values of λ , the curves are bent upward, indicating that the system is active for any λ . This excludes the presence of a finite threshold for diverging N . While Eq. (1) holds for SIS on any graph, Eq. (2) was instead obtained for a specific network model (intrinsically correlated for $\gamma < 3$ and uncorrelated for $\gamma > 3$ [12]). For generic topologies, it is simple to show [11] that $\sqrt{k_c}$ is a lower bound for the largest eigenvalue of the adjacency matrix. This allows to conclude that, unless the degree distribution is strictly bounded from above, the threshold for SIS on any graph vanishes in the thermodynamic limit.

How generic are these results? Prakash et al. [9] have recently argued that Eq. (1) is valid for *all* epidemic processes, regardless of their particular microscopic details. To check this claim, we consider the SIR model. At the HMF level, the threshold takes the value $\lambda_c^{SIR} = \langle k \rangle / [\langle k^2 \rangle - \langle k \rangle]$ [20, 21] and is therefore finite for scale-rich networks with $\gamma > 3$. From the analysis in Eq. (3), on the other hand, it should be vanishingly small in the large network limit, according to Ref. [9]. We have checked this possibility by performing numerical simulations of the SIR model on networks with $\gamma = 4.5$ and different values of N , with fixed $k_{max} = \langle k_{max} \rangle$. In this case, the HMF estimated threshold takes the value $\lambda_c^{SIR} \simeq 0.31$, independent of the network size, while the predictions from Eq. (3) are $\lambda_c^{SIR} \simeq 0.0567, 0.0796$, and 0.1118 for the different network sizes considered. In Fig. 2 we report the final density of infected individuals R as a function of the spreading rate λ , starting from a single randomly chosen infected node. The prediction of HMF

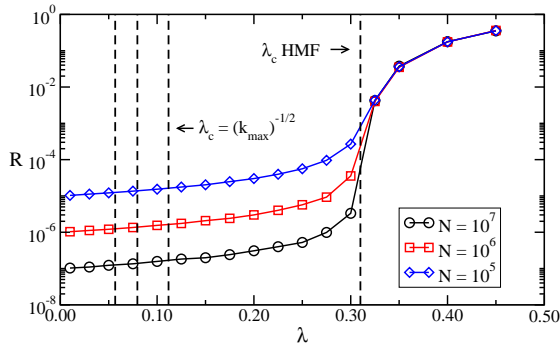


FIG. 2: Total number R of infected individuals in the SIR model on QN of different size N as a function of the spreading rate λ . Networks have $\gamma = 4.5$.

theory seems in this case to be much more accurate than Eq. (3), contrary to the generic claim made in Ref. [9]: The threshold remains finite in the large N limit.

To understand the different behavior of the two models we look at the origin of the incorrect HMF prediction for SIS in QN. From a mathematical point of view, the HMF approach is equivalent [6] to replacing the QN with given adjacency matrix a_{ij} by an annealed network with an averaged adjacency matrix, \bar{a}_{ij} [6, 22]. In the uncorrelated case this matrix reduces to $\bar{a}_{ij} = k_i k_j / [N \langle k \rangle]$, which has a unique nonzero eigenvalue $\bar{\Lambda}_N = \langle k^2 \rangle / \langle k \rangle$. Hence the annealed network approximation destroys the detailed structure of the eigenvalue spectrum of QN and preserves the correct largest eigenvalue only for $\gamma < 5/2$. This basic feature, and not (as suggested in Ref. [16]) the disregard of dynamical correlations, is at root of the inaccuracy of the HMF approach. A more physical insight comes from the analysis of a star graph with one center connected to k_{max} leaves of degree 1. In this case the largest eigenvalue of the adjacency matrix is $\Lambda_N = \sqrt{k_{max}}$ which implies $\lambda_c = 1/\sqrt{k_{max}}$. The same result can be easily recovered by writing the rate equations for the probability ρ_{max} (ρ_1) for the center (leaves) to be active, namely $\dot{\rho}_{max} = -\rho_{max} + (1 - \rho_{max})\rho_1 \lambda k_{max}$ and $\dot{\rho}_1 = -\rho_1 + (1 - \rho_1)\rho_{max} \lambda$. Imposing the steady state condition one finds

$$\rho_{max} = \frac{\lambda^2 k_{max} - 1}{(1 + \lambda k_{max})\lambda} \quad \rho_1 = \frac{\lambda^2 k_{max} - 1}{(1 + \lambda)k_{max}\lambda}, \quad (4)$$

and hence the threshold condition above. The message of Eq. (4) for a generic quenched random graph is strong: Independently from all the rest of the system, for $\lambda > 1/\sqrt{k_{max}}$ the subgraph composed by the node with degree k_{max} and its neighbors is in the active state. This core of activity provides a self-sustained source of infection that, since in the full graph the neighbors of the hub are not leaves, can transfer the activity to their other neighbors and spread in this way the epidemics to a finite fraction of vertices. This is confirmed by Fig. 3,

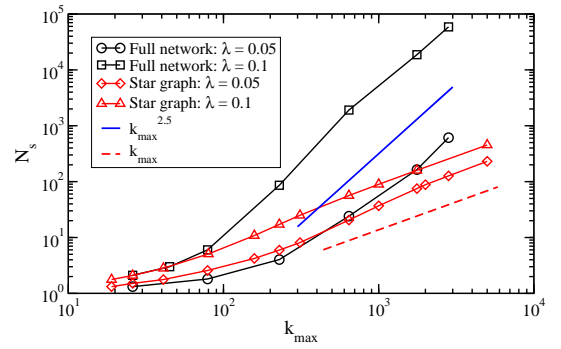


FIG. 3: Number N_s of active nodes in surviving runs for values of λ smaller than the threshold predicted by HMF ($\lambda_c(HMF) = 0.138 \dots$) as a function of k_{max} , compared with the same quantity for star graphs. Networks have $\gamma = 3.5$.

showing the number of active nodes in surviving runs, N_s , on a full network with $\gamma = 3.5$ and on a star graph with the same k_{max} [25]. For $k_{max} < 1/\lambda^2$ the values of N_s in the full network and in the star graph are comparable: both systems are subcritical and the subgraph centered around the node with degree k_{max} is where activity lingers before disappearing. For $k_{max} > 1/\lambda^2$ the star graph becomes active and N_s becomes asymptotically proportional to k_{max} . In the full network instead, the asymptotic behavior is $N_s \sim k_{max}^{\gamma-1} \sim N$, indicating that the active state is endemic: the hub spreads the activity to a finite fraction of the whole system. Reaching the fully endemic state requires larger systems for small λ , but nothing changes qualitatively for any $\lambda > 0$.

Understanding the behavior of SIS allows also to unravel why things go differently for SIR. In the former case, the possibility for hubs to be reinfected multiple times, which allows the presence of a steady state, boosts their impact on the dynamics. In the case of SIR, on the other hand, high-degree vertices can only be infected once and this strongly limits their role in the dynamics. Based on this observation, it is natural to conjecture that epidemic models allowing a steady state, such as SIS, will lead to a null threshold in any infinite QN, while all models without a steady-state will conform with HMF theory, with a finite threshold on scale-rich topologies.

The strong effect of the hub in the dynamics raises further issues on the SIS model. While fixing the value of k_{max} to its ensemble average leads to results consistent with the presence of a non-zero threshold in finite systems, as implied by Eq. (3), if this constraint is relaxed, k_{max} has large sample to sample fluctuations leading to nontrivial consequences. In Fig. 4 we explore the effect of this variability by comparing simulations performed at fixed λ and N , and different values of k_{max} . The growth of the activity density for increasing k_{max} indicates that the relation between the threshold (or the largest eigenvalue) and cut-off k_c , Eq. (3), can in fact

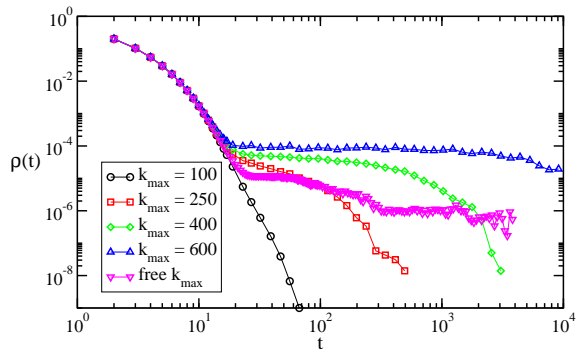


FIG. 4: Decay of the activity density for the SIS model in networks with $\gamma = 4.5$, $N = 10^6$, $\lambda = 0.1$ and changing k_{max} .

be refined, and be expressed in terms of the actual maximum degree, $\lambda_c = 1/\sqrt{k_{max}}$ [18]. However, the large variations of k_{max} among different realizations of the network with the same γ and N do not wash away as N diverges and severely hinder the determination of the threshold in simulations with unrestricted k_{max} . As mentioned before, for $\gamma > 3$ the standard deviation of k_{max} increases as the average value $\langle k_{max} \rangle \sim N^{1/(\gamma-1)}$ [18], and there is always a large sample to sample variability. Hence an unrestricted sampling at fixed λ unwarrantedly averages networks with different thresholds and effective time scales, some subcritical and some supercritical, making impossible even to determine the presence of a well-defined steady state. This fact is exemplified in Fig. 4, where we plot for comparison the activity density obtained averaging over networks with a freely varying k_{max} . For $\gamma < 3$ the situation depends on the way the network is generated, and in particular on the way the upper bound of the degree distribution $M = N^{1/\omega}$ grows [22]. If $\omega = 2$ (uncorrelated configuration model [17]) or larger, the quantity $\langle k^2 \rangle / \langle k \rangle$ becomes sharply peaked as N grows [22]. If instead $\omega = 1$, as in the normal configuration model [23], the ratio $\langle k^2 \rangle / \langle k \rangle$ (and hence the threshold) wildly changes from realization to realization, with relative fluctuations diverging as $N^{2(3-\gamma)(\gamma-2)/(\gamma-1)}$ [22]. Notice that in the intermediate region $5/2 < \gamma < 3$, the average value of $\sqrt{k_{max}}$ is larger than $\langle k^2 \rangle / \langle k \rangle$ but, since fluctuations of the latter diverge, for some network realizations the actual threshold λ_c is much smaller than the value predicted by Eq. (3). We conclude that, unless $\gamma < 3$ and $\omega \geq 2$, no average epidemic threshold can be properly defined from a numerical point of view for networks with unrestricted k_{max} .

In summary, we have studied how the threshold for models of epidemic spreading on quenched scale-rich networks behaves as their size grows. The threshold for SIS model always vanishes in the thermodynamic limit, due to the role of hubs. This bears no relationship, at odds with the predictions of HMF theory, with the divergence

of the second moment of the degree distribution, which is finite. For the SIR model instead the threshold vanishes only for scale-free topologies (either quenched or annealed), in agreement with HMF theory. We conjecture that these different types of behavior are generic for systems possessing (or not) a steady state. While the result of a vanishing threshold for SIS is exact on *quenched* networks, it is however of limited interest from an epidemiological perspective. The interaction patterns over which real diseases spread generally vary over short time scales [24], and are therefore better described by annealed topologies [6], for which HMF theory works by definition, and the threshold is finite for $\gamma > 3$. From a statistical physics point of view, instead, our results open a promising path towards a better understanding of the scope and limits of HMF theory as a theoretical tool to analyze dynamics on heterogeneous networks.

R.P.-S. acknowledges financial support from the Spanish MEC (FEDER), under projects No. FIS2007-66485-C02-01 and FIS2010-21781-C02-01; ICREA Academia, funded by the Generalitat de Catalunya; and the Junta de Andalucía, under project No. P09-FQM4682. We thank A. Vespignani and M. A. Muñoz for discussions.

-
- [1] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
 - [2] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 - [3] M. J. Keeling and K. T. D. Eames, *J. R. Soc. Interface* **2**, 295 (2005).
 - [4] R. M. Anderson and R. M. May, *Infectious diseases in humans* (Oxford University Press, Oxford, 1992).
 - [5] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
 - [6] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).
 - [7] Y. Wang *et al.*, in *22nd International Symposium on Reliable Distributed Systems (SRDS'03)* (2003), pp. 25–34.
 - [8] S. Gómez *et al.*, *Europhys. Lett.* **89**, 38009 (2010).
 - [9] B. A. Prakash *et al.*, *Tech. Rep.* arXiv:1004.0060 (2010).
 - [10] F. Chung, L. Lu, and V. Vu, *Proc. Natl. Acad. Sci. USA* **100**, 6313 (2003).
 - [11] J. G. Restrepo, E. Ott, and B. R. Hunt, *Phys. Rev. E* **76**, 056119 (2007).
 - [12] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *Euro. Phys. J. B* **38**, 205 (2004).
 - [13] A. Ganesh, L. Massoulié, and D. Towsley, in *IEEE INFOCOM* (2005), pp. 1455–1466.
 - [14] M. Krivelevich and B. Sudakov, *Combinatorics, Probability and Computing* **12**, 61 (2003).
 - [15] S. Chatterjee and R. Durrett, *Annals of Probability* **37**, 2332 (2009).
 - [16] R. Durrett, *Proc. Natl. Acad. Sci. USA* **107**, 4491 (2010).
 - [17] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, *Phys. Rev. E* **71**, 027103 (2005).
 - [18] D.-H. Kim and A. Motter, *Phys. Rev. Lett.* **98**, 248701 (2007).

- (2007).
- [19] J. Marro and R. Dickman, *Nonequilibrium phase transitions in lattice models* (Cambridge University Press, Cambridge, 1999).
 - [20] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000).
 - [21] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).
 - [22] M. Boguñá, C. Castellano, and R. Pastor-Satorras, Phys. Rev. E **79**, 036110 (2009).
 - [23] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
 - [24] C. T. Butts, Science **325**, 414 (2009).
 - [25] Analogous results are obtained for different values of γ (data not shown)